Australian Government
Department of Defence
Defence Science and
Technology Organisation

Computing with Epistemic Uncertainty

Lewis Warren

National Security and ISR Division
Defence Science and Technology Organisation

DSTO-TN-1394

ABSTRACT

This report proposes the use of uncertainty management principles for processing combinations of aleatory and epistemic uncertainty forms through arithmetic operations to yield an output uncertainty margin as an extreme value interval. The method is demonstrated using some test equations from an Epistemic Uncertainty Workshop conducted in 2002 by the Sandia National Laboratories in the USA. The approach is robust, computationally efficient and does not require special assumptions. A comparison with the results of the workshop participants showed similar results to most workshop participants. The benefit of this approach is that it is a step-wise process for computation with no need for a large number of simulations nor complex sampling strategies.

RELEASE LIMITATION

Approved for public release

Published by

National Security and ISR Division
DSTO Defence Science and Technology Organisation
PO Box 1500
Edinburgh South Australia 5111 Australia

Telephone: 1300 333 362
Fax: (08) 7389 6567

AR-016-210
January 2015

APPROVED FOR PUBLIC RELEASE

Computing with Epistemic Uncertainty

Executive Summary

In computational models of real-world problems input measures for elements about which little data exists are sometimes subjectively estimated. Such input estimates are said to embed epistemic uncertainty due to the lack of knowledge about the elements. The presence of such epistemic uncertainty, especially in combination with probabilistic aleatory uncertainty, is known to present theoretical challenges when propagating uncertainty through mathematical equations in complex system models. We demonstrate the use of uncertainty management principles for processing mixed uncertainty forms through arithmetic operations to determine an output margin as an extreme value interval. In August 2002 Sandia National Laboratories in the US conducted an Epistemic Uncertainty Workshop where the participants were invited to quantify the amount of uncertainty in the outputs of two relatively simple equations, and the inputs contained both epistemic and aleatory uncertainty forms. The proposed method is applied to the Sandia Workshop problems and our results are compared with those of other approaches that have also been published for those problems. We conclude that our results are similar to those obtained by most workshop participants, and furthermore, that our approach lends itself to a stepwise process without the need for complex simulations as was the chosen approach by many researchers who have addressed those problems.

This page is intentionally blank

Contents

This page is intentionally blank

1. Introduction

The presence of epistemic uncertainty, especially in combination with probabilistic aleatory uncertainty, presents a theoretical challenge when propagating uncertainty through mathematical equations. Towards finding improved methods and techniques for propagating hybrid uncertainty forms through system equations, Sandia National Laboratories in the USA have researched many theoretical approaches to the problem in relation to reliability and safety evaluation of nuclear products and facilities, as by Diegert *et al*. (2007).

In 2002 Sandia conducted an Epistemic Uncertainty Workshop, described by Oberkampf *et al*. (2002), organised around a set of challenge problems where the participants were invited to quantify the amount of uncertainty in the outputs of two relatively simple equations, where the inputs contained both epistemic and probabilistic aleatory uncertainty. A summary of the workshop results by Ferson *et al.* (2004) indicated that a diverse range of approaches were applied.

Since that workshop Sandia and others have conducted further exploratory analysis of alternative methods to propagate hybrid mixtures of uncertainty through mathematical equations, many of which applied simulation with various sampling regimes, e.g. Ferson and Hajagos (2004), Tonon (2004), Helton *et al*. (2004), Diegert *et al.* (2007), Rao *et al.* (2007), Eldred and Swiler (2009) and Helton and Johnson (2011). Many authors who applied simulation techniques were also required to invoke the Laplace principle of insufficient reason and assume a uniform probability distribution across the input epistemic intervals for sampling purposes.

As an alternative to simulation and the complex computations required in the methods above, we propose a simple approach that avoids the complications and assumptions of other approaches while still yielding realistic outputs. As described by Helton and Johnson (2011), the stated objective of this hybrid uncertainty research at Sandia was to determine the margin of uncertainty in output values, and margin means range. Thus we focus only on the extreme points that will determine this range of output values of the test functions. The proposed approach to determine these extreme values is founded on some general uncertainty management principles combined with the application of interval arithmetic, as in Moore (1966), and standard function optimisation techniques using MATLAB. The only assumption underlying this approach is that any value that is explicit or implicit in the input information is possible.

The structure of the paper is as follows. Section 2 describes the two broad classes of uncertainty and the general uncertainty management principles upon which the proposed approach is based. Section 3 introduces the Sandia workshop problems and the steps in applying the proposed approach. Section 4 then details the computations for the Sandia problems and Section 5 compares those results with other published results. Finally, Section 6 offers some conclusions.

2. Management of Hybrid Uncertainty

The various forms of uncertainty in the input data used by computational models are often simplified into two main classes: aleatory and epistemic uncertainty. These may be present separately or in conjunction in various proportions which then comprise hybrid uncertainty combinations. First, we will outline descriptions of each main class from the Sandia workshop organisers, and subsequently, some general uncertainty management principles that support our approach.

2.1 Aleatory uncertainty

As described by Oberkampf *et al*. (2002):

We use the term aleatory uncertainty to describe the inherent variation associated with the physical system. …The mathematical representation most commonly used for aleatory uncertainty is a probability distribution. When substantial experimental data are available for estimating a distribution, there is no debate that the correct model for aleatory uncertainty is a probability distribution.

In other words, this type of uncertainty concerns measure or value variations that are governed by a random process and computations are guided by well developed probabilistic methods. Aleatory uncertainty is sometimes called irreducible uncertainty because it originates from uncontrollable stochastic variations.

2.2 Epistemic uncertainty

Again, as described by Oberkampf *et al*. (2002):

Epistemic uncertainty derives from some level of ignorance of the system or the environment. We use the term epistemic uncertainty to describe any lack of knowledge or information in any phase or activity of the modelling process. The key feature that this definition stresses is that the fundamental cause is incomplete information or incomplete knowledge of some characteristic of the system or environment.

A lack of knowledge may have many causes but some examples may include: sparse evidence or data, lack of experience, incomplete understanding of the context or environment, a non-stationary context, a partial understanding of system elements, or unknown or even unknowable system parameter values. In the data for the Sandia workshop problems many estimates of the input parameter values are given as epistemic intervals of possible minimum and maximum values. Epistemic uncertainty is sometimes called reducible uncertainty, as it may be reduced with the supply of more data or information.

2.3 Uncertainty management principles

In the proposed computational approach, the guidance for each step will be provided by three uncertainty management principles as have been outlined by Klir (1990,1995). Although these are very simple, they can have a strong supervisory effect on computations.

2.3.1 Principle of maximum uncertainty

The principle of maximum uncertainty (PMxU) supervises how different forms of input uncertainty should be combined into a uniform representation reflecting the highest form of input uncertainty. Using that uniform representation, the maximum uncertainty in the inputs is propagated through computations and hence is truly reflected in the output uncertainty levels. Thus, when epistemic interval measures are to be combined with probabilistic data, the result should be an epistemic interval which may also embed some aleatory information. One example of a violation of this principle is when a uniform probability distribution is assumed to apply within an epistemic interval. Such an assumption adds information and consequently reduces the uncertainty in the input data.

2.3.2 Principle of minimum uncertainty

The principle of minimum uncertainty (PMnU) dictates that if alternative types of investigative analysis or information reduction techniques are applied to the source data, no new uncertainty should be added by the method adopted. An example of how this principle is violated is when a particular sampling strategy is applied in simulation. The reason for this is that there is an additional uncertainty associated with each sampling method since different sampling strategies may produce different results.

2.3.3 Principle of uncertainty invariance

The principle of uncertainty invariance (PInvU) dictates that no extra uncertainty is added or subtracted when transforming a quantity between the representations used by different uncertainty formalisms. The previous example of transforming an epistemic interval into a uniform probability distribution between the extremities violates this principle also because it adds information and thus reduces the original uncertainty.

3. The Workshop Problems Overview

The workshop presented two problems being two relatively simple functions as shown in Section 5. Applied to the first expression were twelve different sets of input data with hybrid uncertainty forms. The second problem concerned a mass-spring-damper system with a periodic force applied to the mass. An expression for the steady state magnification factor of the system was given based on the ratio of the mass lateral displacement and the maximum input force (and a spring constant). Only one input dataset with hybrid uncertainties was applied in the second problem.

4. The Proposed Approach

4.1 The interval arithmetic operations

The required interval operations, as described by Moore (1966), are as follows. In other cases, such as for exponentiation, the output interval is defined by the minimum across all combinations of the input interval values and the exponent interval values, and similarly the maximum across all combinations of those values.

$$[a,b]+[c,d] = [a+c, b+d]$$

$$[a,b]-[c,d] = [a-d, b-c]$$

$$[a,b]\times[c,d]$$
$$= [\min(a\times c, a\times d, b\times c, b\times d), \max(a\times c, a\times d, b\times c, b\times d)]$$

$$[a,b]\div[c,d]$$
$$= [\min(a\div c, a\div d, b\div c, b\div d), \max(a\div c, a\div d, b\div c, b\div d)]$$
$$\textit{when } 0 \textit{ is not in } [c,d]$$

4.2 The computation steps

Since the epistemic input uncertainties are axiomatically a higher uncertainty form than the aleatory inputs, in accordance with the PMxU we use epistemic intervals as the medium by which we synthesise the input aleatory and epistemic information. Thus, there are three steps in the proposed approach. First, we determine the minimum and maximum parameter values that are possible from the epistemic data and aleatory data (at given confidence degrees where such are required). Next, we apply these input parameter extreme value intervals to the equations to determine a conservative range of output values for each function by interval arithmetic. The third step is then to determine the true output minimum and maximum values for the functions which exist within this conservative epistemic interval. To do this we apply standard methods and tools for non-linear function optimisation subject to parameter constraints.

4.3 Synthesis of the hybrid uncertainties

The input data for the parameters in both challenge problems consists of a mixture of epistemic intervals, lognormal probability distributions, crisp triangular probability distributions (TPD), and fuzzy TPD where the three vertices are given as crisp epistemic intervals. The conversion of the input parameter information to epistemic intervals is as follows, with the parameter epistemic intervals subsequently processed through the equations by interval arithmetic.

4.3.1 Converting the lognormal distributions to epistemic intervals

For the lognormal parameter probability distributions a confidence level is applied to determine statistical confidence intervals for parameter values. And since aleatory information has less uncertainty than epistemic inputs, we can consider these limits as epistemic intervals where all values within the range are possible at the given confidence level.

4.3.2 Converting the crisp TPDs to epistemic intervals

The traditional type of confidence interval referred to above is associated with a given degree of probabilistic confidence. For a TPD the mean ± 2 standard deviations would then yield a 98% degree of confidence. However, unlike many other probability distributions the TPD has a discrete base which is the maximum range of values that is possible. Hence, by the PMxU the interval conversion for a crisp TPD is the size of the base of the TPD.

4.3.3 Converting the fuzzy TPDs to epistemic intervals

Again, to conform to the PMxU in these uncertainty transforms, the values from the input intervals for the base points determine the most extreme values and are selected as follows.

The extreme base points:
Min of left base point interval = a
Max of right base point interval = b
Then the final epistemic interval is simply [a, b].

5. The Problem Computations

The solutions for Problem 1 data sets (i-xii) and Problem 2 by applying the uncertainty management principles with interval algebra and optimisation methods are as follows.

5.1 Problem 1 computations

$$\textbf{For} \quad y = (a+b)^a . \tag{1}$$

(i) a = [0.1, 1.0], b = [0, 1.0]

$$y = [(0.1+0),(1.0+1.0)]^{[0.1,1.0]}$$
$$= [0.1\,,\, 2.0]^{[0.1,1.0]}$$

Lowest exponential value = $0.1^{1.0}$ = 0.1
Highest exponential value = $2.0^{1.0}$ = 2.0
i.e. y = [0.1, 2.0] is the epistemic constraint interval.

However, 0.1 is not necessarily a true function minimum. The true minimum is found by taking the partial derivatives w.r.t. a and b and letting both equal zero. Since $a \geq 0.1$, the

min must be when $b = 0$. Then the turning point of $y = a^a$ is at $a = e^{-1}$ and $y = 0.6922$. And since the function increases monotonically with b above this point, the maximum is *2.0* at the maximum of the epistemic constraint interval.

Thus, the true extreme value interval is y = [0.69, 2.0].

(ii) a = [0.1, 1.0]
b = [0.6, 0.8], [0.4, 0.85], [0.2, 0.9],
[0.0, 1.0]
All equally credible so by PMxU:
b = [Min , Max] = [0.0 , 1.0]
Thus a, b same as (i) and y = [0.69, 2.0]

(iii)a = [0.1, 1.0]
b = [0.6, 0.0], [0.4, 0.8], [0.1, 0.7],
[0.0, 1.0]
Thus Min and Max same as (ii) and b = [0.0 , 1.0]
Again y = [0.69, 2.0]

(iv) a = [0.1, 1.0]
b = [0.6, 0.8], [0.5, 0.7], [0.1, 0.4],
[0.0, 1.0]
Thus Min and Max same as (ii) and b = [0.0 , 1.0]
Again y = [0.69, 2.0]

(v) a = [0.5, 0.7], [0.3, 0.8], [0.1, 1.0]
b = [0.6, 0.6], [0.4, 0.85], [0.2, 0.9],
[0.0, 1.0]
a [Min, Max] = [0.1 , 1.0]
b [Min, Max] = [0.0 , 1.0]
Again y = [0.69, 2.0]

(vi) a = [0.5, 1.0], [0.2, 0.7], [0.1, 0.6]
b = [0.6, 0.6], [0.4, 0.8], [0.1, 0.7],
[0.0, 1.0]
a [Min, Max] = [0.1 , 1.0]
b [Min, Max] = [0.0 , 1.0]
Again y = [0.69, 2.0]

(vii) a = [0.8, 1.0], [0.5, 0.7], [0.1, 0.4]
b = [0.8, 1.0], [0.5, 0.7], [0.1, 0.4],
[0.0, 0.2]
a [Min, Max] = [0.1 , 1.0]
b [Min, Max] = [0.0 , 1.0]
Again y = [0.69, 2.0]

(viii) a = [0.1, 1.0]
and b is given by a lognormal probability distribution where ln b ~ N (μ, σ)

and μ = [0.0, 1.0] and σ = [0.1, 0.5].
For Normal distribution 99.7% Confidence interval, CI = $\mu \pm 3\sigma$

Thus by PMxU:

$$
\begin{aligned}
CI\,(Min) &= \mu\,(Min) - 3\,\sigma\,(Max) \\
&= 0 - 3\,(0.5) \\
&= -1.5 \\
CI\,(Max) &= \mu\,(Max) + 3\,\sigma\,(Max) \\
&= 1 + 3\,(0.5) \\
&= 2.5 \\
\\
CI\,(X) &= [\,-1.5,\ 2.5\,]
\end{aligned}
$$

For lognormal distribution, $b = e^X$
Thus, $b\,(Max) = e^{2.5} = 12.182$
$b\,(Min) = e^{-1.5} = 0.223$
i.e. b = [0.223, 12.182]

Then, $y = [\,a + b\,]^a$

$$
\begin{aligned}
&= [[0.1, 1.0] + [0.223, 12.182]]^{[0.1\,,\,1.0]} \\
&= [0.323, 13.182]^{[0.1\,,\,1.0]} \\
&= [\,Min, Max\,] \\
&= [\,0.323,\ 13.182\,]
\end{aligned}
$$

But again this lower interval boundary is not a true value for the function and the true lower boundary is the same as in the previous cases. However, the upper boundary is a true value due to the increased upper boundary of the b parameter for the monotonically increasing function.

Thus, y = [0.69, 13.18]

(ix) a = [0.5, 0.7], [0.3, 0.8], [0.1, 1.0]
and b is given by a lognormal probability distribution where ln b ~ N (μ, σ)
and μ = [0.6, 0.8], [0.2, 0.9], [0.0, 1.0]
and σ = [0.3, 0.4], [0.2, 0.45], [0.1, 0.5]

By PMxU, a = [0.1, 1.0],
μ = [0.0, 1.0], σ = [0.1, 0.5]
Values for a, μ and σ are same as (viii) and,
y = [0.69, 13.18]

(x) a = [0.5, 1.0], [0.2, 078], [0.1, 0.6]
and b is given by a lognormal probability distribution where ln b ~ N (μ, σ)
and μ = [0.6, 0.9] , [0.1, 0.7] , [0.0, 1.0]
and σ = [0.3, 0.45], [0.15, 0.35], [0.1, 0.5]

By PMxU , a = [0.1, 1.0],
μ = [0.0, 1.0, and σ = [0.1, 0.5]
Values for a, μ and σ are same as (viii) and,

$y = [0.69,\ 13.18]$

(xi) $a = [0.8, 1.0], [0.5, 0.7], [0.1, 0.4]$
and b is given by a lognormal probability distribution where $\ln b \sim N(\mu, \sigma)$
and $\mu = [0.6, 0.8], [0.1, 0.4], [0.0, 1.0]$
and $\sigma = [0.4, 0.5], [0.25, 0.35], [0.1, 0.2]$

By PMxU, $a = [0.1,\ 1.0]$,
$\mu = [0.0, 1.0]$ and $\sigma = [0.1, 0.5]$
Values for a, μ and σ are same as (viii) and,
$y = [0.69,\ 13.18]$

(xii) $a = [0.1, 1.0]$
b is given by a lognormal probability distribution where $\ln b \sim N(\mu, \sigma)$
and $\mu = 0.5$ and $\sigma = 0.5$.
For Normal distribution 99.7% Confidence Interval, $CI = \mu \pm 3\sigma$

Thus by PMxU:

$$
\begin{aligned}
CI(Min) &= \mu(Min) - 3\sigma(Max) \\
&= 0.5 - 3(0.5) \\
&= -1.0 \\
CI(Max) &= \mu(Max) + 3\sigma(Max) \\
&= 0.5 + 3(0.5) \\
&= 2.0 \\
CI(X) &= [-1.0,\ 2.0]
\end{aligned}
$$

And for lognormal distribution, $b = e^{X}$
Thus, $b(max) = e^{2.0} = 7.389$
$b(min) = e^{-1.0} = 0.368$
i.e. $b = [0.368,\ 7.389]$

$$
\begin{aligned}
\text{Then, } y &= [a + b]^{a} \\
&= [[0.1, 1.0] + [0.368, 7.389]]^{[0.1, 1.0]} \\
&= [0.468,\ 8.389]^{[0.1, 1.0]} \\
&= [Min, Max] \\
&= [0.468,\ 8.389] \text{ at 99.7\% confidence level.}
\end{aligned}
$$

Again this lower interval boundary is not a true value for the function and the true lower boundary is the same as in the previous cases. However, the upper boundary is a true value due to the increased upper boundary of the b parameter for the monotonically increasing function.

Thus, $y = [0.69,\ 8.39]$

Caveat for the problems with lognormal data:

Because the Normal distribution is asymptotic to the base and very small probabilities may occur over a large value range, 99.7% confidence limits are associated with the results for the lognormal data, unlike the problems with TPD or epistemic intervals which reflect 100% confidence in the output interval. Thus there is still a small chance of a value beyond the output intervals for the lognormal data examples.

5.2 The Problem 2 Computations

A simple linear oscillator system consists of a mass, spring and a damper with a periodic forcing function applied to the mass. The steady state magnification factor (D_S) is given by the following expression as provided. The problem is to estimate the uncertainty in the estimate for D_S.

$$D_S = \frac{X}{Y/k} \qquad (2)$$

and

$$D_S = \frac{k}{\sqrt{\left(k - m\omega^2\right)^2 + (c\omega)^2}} \qquad (3)$$

…where, m = mass of cart, k = spring constant, c = linear damping coefficient, Y and ω = the amplitude and frequency of the sinusoidal excitation (Ycos ωt), and X = lateral displacement.

As previously described, all inputs will first be reduced to intervals by the PMxU, and initially we estimate a conservative output interval by using interval arithmetic. The true minimum and maximum values within the output epistemic conservative interval are then determined by optimising a non-linear function with input parameters subject to constraints.

The Input parameter data:

m: a crisp triangular probability distribution,
(min, mod, max) = (10, 11, 12)

k: Three independent sources estimate a fuzzy triangular probability distribution:

Min =	$[a_1, a_2]$,	Mod = $[h_1, h_2]$,	Max = $[b_1, b_2]$
1:	[90, 100],	[150, 160],	[200,210]
2	[80, 110],	[140, 170],	[200,220]
3:	[60, 120],	[120, 180],	[190,230]

c: Three independent interval estimates for
[cmin, cmax], 1: [5, 10], 2: [15, 20], 3: [25, 25]

Y : a single interval estimate: [Ymin, Ymax]
= [70, 100]

ω: a single fuzzy triangular probability distribution :

Min = $[a_1, a_2]$,	Mod = $[h_1, h_2]$,	Max = $[b_1, b_2]$,
[2.0, 2.3],	[2.5, 2.7],	[3.0, 3.5].

5.2.1 Converting problem 2 data to epistemic intervals

As in Problem 1, the input data consists of epistemic intervals, and crisp or fuzzy triangular probability distributions. Again, all input uncertainty representations will be reduced to epistemic intervals according to the PMxU.

The m data:

For the TPD, by the PMxU the extreme value interval is [10, 12].

The k data:

From the 3 estimates the lowest Min = 60

From the 3 estimates the largest max = 230

By the PMxU the k interval with maximum value uncertainty is [60, 230].

The c data:

From the 3 interval estimates, by the PMxU the combined interval is [5, 25].

The Y data : Input is a single interval [70, 100].

The ω data:

For the single fuzzy TPD the lowest base minimum in that interval estimate is 2.0 and the highest base maximum from that interval estimate is 3.5. Thus, the epistemic interval embedding the maximum value uncertainty is the base interval [2.0, 3.5].

5.2.2 The interval arithmetic operations

Interval arithmetic is next applied to determine the large conservative output interval within which the true function extreme values must be found.

The problem function is:

$$D_S = \frac{k}{\sqrt{\left(k - m\omega^2\right)^2 + \left(c\omega\right)^2}} \tag{3}$$

$$\begin{aligned}(c\omega)^2 &= [\ [5, 25] \,.\, [2, 3.5]\ ]^2 \\ &= [\ 10, 87.5\ ]^2 \\ &= [\ 100.00, 7656.25\ ]\end{aligned}$$

$$\begin{aligned}m\omega^2 &= [\ 10, 12\ ] \,.\, [\ 2, 3.5\ ]^2 \\ &= [\ 10, 12\ ] \,.\, [\ 4, 12.25\ ] \\ &= [\ 40.00, 147.00\ ]\end{aligned}$$

$$\begin{aligned}k - m\omega^2 &= [\ 60.00, 230.00\ ] - [\ 40.00, 147.00\ ] \\ &= [\ (60-147), (230-40)\ ] \\ &= [\ -87.00, 190.00\ ]\end{aligned}$$

$$\left(k - m\omega^2\right)^2 = [\,7569,\ 36100.00\,]$$

$$\left(k - m\omega^2\right)^2 + \left(c\omega\right)^2 = [\,7569,\ 36100.00\,] + [\,100.00,\ 7656.25\,]$$
$$= [\,7669.00,\ 43756.25\,]$$

$$\sqrt{\left(k - m\omega^2\right)^2 + \left(c\omega\right)^2} = [\,7669.00,\ 43756.25\,]^{0.5}$$
$$= [\,87.60,\ 209.18\,]$$

Then inserting for D_S,

$$D_S = \frac{[\,60.00,\ 230.00\,]}{[\,87.60,\ 209.18\,]}$$
$$= \left[\frac{60.00}{209.18},\ \frac{230.00}{87.60}\right]$$
$$= [\ 0.29,\ \ 2.63\ ]$$

This is the conservative estimate of the possible output extreme values within which the true function minimum and maximum values must be found.

5.2.3 The non-linear function optimisation

The true minimum and maximum within the possible output D_S interval is determined by following standard methods for the optimisation of non-linear functions of several variables under constraints. Using MATLAB mathematical software the function was found to be monotonically increasing within the D_S interval constraint so the maximum within that PMxU output interval is 2.63. If the surface is cut at the value of 2.63 there may also be more than one parameter set to yield that value. One such set is shown below.

Maximum D_S = 2.63:
when k = 109.41, m = 12, c = 5 w = 3.5
(Occurs when 3 parameters are at input boundary points.)

Minimum D_S = 0.49:
when k = 60, m = 12, c = 25, w = 3.5
(Occurs when all parameters are at input boundary points.)

Thus, the output extreme values are DS = [0.49, 2.63].

6. Comparison with Other Approaches

Some published results of workshop participants as summarised in Ferson *et al* (2004) and Ferson and Hajagos (2004), plus some subsequent explorations of alternative approaches by Helton *et al.* (2004), are compared with those of this author as shown in Tables 1 and 2. Table 1 lists the approximate maximum range across the results of the workshop

participants for the output mean value range, as well as the output extreme value range for Problem 1. The approximate extreme values for the results of later exploratory investigations by Helton *et al.* (2004) were taken from the published cumulative distribution functions (CDF) for those investigations. If we then assume the upper extreme value is the most critical for each estimate of the margin of output uncertainty, the last column in the table shows where a significant difference occurs (and the approximate % difference) between the results of the proposed approach and the upper result of the various published investigations. Considering the maximum values that could be found in the published results, for Problem 1 there was no difference in half of the data sets (the simpler ones), and a 20-45% increase in the other data sets.

Table 1: Summary of published results for Problem 1

Dataset Number Original	Current	Workshop Participant Results: Mean Values [2, Table 1]	Workshop Participant Results: Extreme Values [3, Figs.4,6,7,15]	Alternative Explorations Results: Extreme Values [5, Figs.15-19]	Proposed Method Results: Extreme Values	Difference in Maximum of Extreme Values
1	i	[0.69, 2.00]	[0.69, 2.00]	[0.69, 2.0]	[0.69, 2.0]	No
2a	ii	[0.84, 1.89]	[0.69, 2.00]	[0.69, 2.0]	[0.69, 2.0]	No
2b	iii	[0.82, 1.85]	[0.69, 2.00]	[0.69, 2.0]	[0.69, 2.0]	No
2c	iv	[0.69, 2.00]	[0.69, 2.00]	[0.69, 2.0]	[0.69, 2.0]	No
3a	v	[0.83, 1.56]	[0.69, 2.00]	[0.69, 2.0]	[0.69, 2.0]	No
3b	vi	[0.82, 1.44]	[0.69, 2.00]	[0.69, 2.0]	[0.69, 2.0]	No
3c	vii	[0.69, 2.00]	[0.69, 2.00]	[0.69, 2.0]	[0.69, 2.0]	No
4	viii	[0.86, 4.42]	[1.0, 11]	[1, 5]	[0.69, 13.2]	Yes (+20%)
5a	ix	[1.05, 3.79]	[0.91, 10.9]	[1, 5]	[0.69, 13.2]	Yes (+21%)
5b	x	[1.03, 3.48]	[0.91, 10.9]	[1, 5]	[0.69, 13.2]	Yes (+21%)
5c	xi	[1.01, 4.08]	[0.94, 9.10]	[1, 5]	[0.69, 13.2]	Yes (+45%)
6	xii	[1.02, 2.89]	[1.20, 7.00]	[1, 5]	[0.69, 8.39]	Yes (+20%)

Table 2 next shows the mean and upper extreme value estimates of those authors who published results for Problem 2. The upper extreme value of 2.63 of the proposed method is approximately equal to the mean estimates of Tonon (2.13) and the maximum mean estimate of Helton *et al.* (2.8). However, our upper output value is significantly less than

the upper limits of Helton *et al.* (~4.0), Ferson and Hajagos (~8.0) and Tonon (~6), which occur at very low probabilities in the tails of their respective CDF of output values.

Notably, with the proposed approach all the aleatory input information is not always required since epistemic uncertainty is a higher uncertainty form. For example, in workshop Problem 2 the probabilistic information within the triangular distribution boundaries was not needed to determine the extremities of the output D_S epistemic interval. However, probability distribution parameters (mean and standard deviation) are used to derive statistical confidence intervals in some Problem 1 examples, and hence the output extreme values are associated with a certain confidence level in those examples.

Table 2: Summary of published results for Problem 2

	Tonon [4]	Ferson and Hajagos [3]	Helton *et al.* [5]	Proposed Method
Mean estimates	2.13	3.72	2.8	NA
Upper extreme value	~ 6	~ 8	~ 4	2.63

7. Conclusions

A computationally simple method founded on uncertainty management principles has been proposed for synthesising and propagating hybrid uncertainties through mathematical equations. From our results we suggest that the application of uncertainty management principles to guide the synthesis and propagation of epistemic and aleatory uncertainty is a computationally efficient approach for determining output extreme values when propagating such hybrid uncertainty combinations through arithmetic operations. The overall effect of applying the uncertainty management principles is that no special techniques or assumptions have been applied that modified the input uncertainties in any way. And by avoiding the need for simulation, various assumptions and selection of specific sampling strategies that may affect results are also avoided. According with the Principle of Maximum Uncertainty, epistemic intervals represent the highest input uncertainty form and thus have been adopted as the single uncertainty representation to synthesise and propagate the hybrid input uncertainties and yield a potential extreme value range. The true function extreme values within that range were then determined by function optimisation techniques. Using two standard test equations, the results of applying the proposed method to compute extreme values were then compared with some other published approaches. One benefit of combining interval arithmetic with the Principle of Maximum Uncertainty is that the limits beyond which output values are impossible are clearly defined. This can then facilitate the interpretation of output cumulative probability distributions when simulation methods are applied. However, we consider the main benefit of the proposed approach is computational efficiency due to a discrete programmable process without the need for complex simulation procedures.

8. References

1. Oberkampf, W., Helton, J., Joslyn C, Wojtkiewicz, S. and Ferson, S.: Challenge Problems (2002): Uncertainty in System Response Given Uncertain Parameters. http://sandia.gov/epistemic, Accessed May 2011.
2. Ferson, S., Joslyn, C., Helton, J., Oberkampf, W. and Sentz, K. (2004): Summary from the epistemic uncertainty workshop: consensus and diversity, *Reliability Engineering and System Safety*. **85:** 355-369.
3. Ferson, S. and Hajagos, J. (2004): Arithmetic with uncertain numbers: rigorous and (often) best possible answers, *Reliability Engineering and System Safety*. **85:** 135-152.
4. Tonon, F. (2004): Using random set theory to propagate epistemic uncertainty through a mechanical system. *Reliability Engineering and System Safety*. **85:** 169-181.
5. Helton, J., Johnson, D. and Oberkampf, W. (2004): An exploration of alternative approaches to the representation of uncertainty in model predictions, *Reliability Engineering and System Safety*. **85:** 39-71.
6. Diegert, K., Klenke, S., Novotny, G., Paulsen, R., Pilch, M. and Trucano, T. (2007): Towards More Rigorous Application of Margins and Uncertainties within the Nuclear Weapons Life Cycle: A Sandia Perspective. *SAND2007-6219*, Sandia National Laboratories, Albuquerque, NM.
7. Rao, K., Kushwaha, H., Verma, A. and Srividya A. (2007): Epistemic uncertainty propagation in reliability assessment of complex systems. *Int. J. of Performability Engineering*. **3**(4): 71-84.
8. Eldred, M. and Swiler, L. (2009): Efficient Algorithms for Mixed Aleatory-Epistemic Uncertainty Quantification with Application to Radiation-hardened Electronics. *SAND2009-5805*, Sandia National Laboratories, Albuquerque, NM.
9. Helton, J. and Johnson, J. (2011): Quantification of margins and uncertainties: Alternative representations of epistemic uncertainty. *Reliability Engineering and System Safety*. **96:** 1034-1052.
10. Moore, R. (1966): *Interval Analysis.* Prentice-Hall, Englewood Cliffs N.J.
11. Klir, G. (1990): A principle of uncertainty and information invariance. *Int. J. General Systems*. **17:** 249-275.
12. Klir, G. (1995): Principles of uncertainty: What are they? Do we need them? *Fuzzy Sets and Systems*. **74:** 15-31.

Page classification: UNCLASSIFIED

DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION DOCUMENT CONTROL DATA	1. DLM/CAVEAT (OF DOCUMENT)

2. TITLE	3. SECURITY CLASSIFICATION (FOR UNCLASSIFIED REPORTS THAT ARE LIMITED RELEASE USE (L) NEXT TO DOCUMENT CLASSIFICATION)
Computing with Epistemic Uncertainty	Document (U) Title (U) Abstract (U)
4. AUTHOR(S)	5. CORPORATE AUTHOR
Lewis Warren	DSTO Defence Science and Technology Organisation PO Box 1500 Edinburgh South Australia 5111 Australia

6a. DSTO NUMBER	6b. AR NUMBER	6c. TYPE OF REPORT	7. DOCUMENT DATE
DSTO-TN-1394	AR-016-210	Technical Note	January 2015

8. FILE NUMBER	9. TASK NUMBER	10. TASK SPONSOR	11. NO. OF PAGES	12. NO. OF REFERENCES
2014/1193838/1	NS 07/418	Head Language Technologies & Fusion (HLTF)	14	12

13. DSTO Publications Repository	14. RELEASE AUTHORITY
http://dspace.dsto.defence.gov.au/dspace/	Chief, National Security and ISR Division

15. SECONDARY RELEASE STATEMENT OF THIS DOCUMENT

Approved for public release

OVERSEAS ENQUIRIES OUTSIDE STATED LIMITATIONS SHOULD BE REFERRED THROUGH DOCUMENT EXCHANGE, PO BOX 1500, EDINBURGH, SA 5111

16. DELIBERATE ANNOUNCEMENT

No Limitations

17. CITATION IN OTHER DOCUMENTS Yes

18. DSTO RESEARCH LIBRARY THESAURUS

Epistemic uncertainty, Hybrid uncertainty, Uncertainty management principles, Extreme values

19. ABSTRACT

This report proposes the use of uncertainty management principles for processing combinations of aleatory and epistemic uncertainty forms through arithmetic operations to yield an output uncertainty margin as an extreme value interval. The method is demonstrated using some test equations from an Epistemic Uncertainty Workshop conducted in 2002 by the Sandia National Laboratories in the USA. The approach is robust, computationally efficient and does not require special assumptions. A comparison with the results of the workshop participants showed similar results to most workshop participants. The benefit of this approach is that it is a step-wise process for computation with no need for a large number of simulations and complex sampling strategies.

Page classification: UNCLASSIFIED